

[illegible]

(i) APPLICANT:

- (A) NAME: NOVARTIS AG
- (B) STREET: Schwarzwaldallee 215
- (C) CITY: Basel
- (E) COUNTRY: Switzerland
- (F) POSTAL CODE (ZIP): 4058
- (G) TELEPHONE: +41 61 69 11 11
- (H) TELEFAX: + 41 61 696 79 76
- (I) TELEX: 962 991

(ii) TITLE OF INVENTION: Improvements in or relating to organic compounds

iii) NUMBER OF SEQUENCES: 33

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

- (ii) TITLE OF INVENTION: Improvements in or relating to organic compounds

(iv) COMPUTER READABLE FORM:

- (2) INFORMATION FOR SEQ ID NO: 1:

(A) LENGTH: 6695 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

- (A) LENGTH: 6695 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Daucus carota*

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 3696..6617

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 3731..3802

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 3851..3979

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4124..4211

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4284..4357

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4430..4528

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4642..4757

00180700 11500
00011 00208160

(ix) **FEATURE:**

(A) NAME/KEY: intron

(B) LOCATION: 5295..5803

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

(B) LOCATION: 6197..6339

TCTAGATGAC	GAAATCGCGC	TACCTTTGAT	TINGAAATAC	TAGGTTGTAG	TATCTTGATT	60
AGTTTTTTTGG	ATATCTTGCT	GTAATTTCTT	TAGGAGATGC	AAACGGTCTT	CATTTAATAT	120
GAGCCCTTGT	GACTTGACAA	AAGTATCTAG	CATGTTTGAT	CACGAGGTAG	CTAAAAAGTA	180
GCGTGTTTGA	TTAAGCACAT	AATATTGTAT	TGGGCCTATT	GGCTATCAAT	GAAGTTTGAT	240
GCAAGTATAT	AGCTTGTATT	ATGCATGTGA	TGAGGGTATA	TAAAAGAAGT	AAAGAACAAT	300
CTCTCGTAGC	ATTCATTTTT	CTCTTGCCTA	TAGTTAACGA	GTTTTGTAC	ACATGACGTT	360
GAAACTGGAT	GTGTCTGTTT	TTCCATCTAA	GTTTGGATTA	CCTGATAGAT	GCTCAACTTC	420
TTGTCAGCC	TTTTCTTTCC	GATTTTTTCC	AAGACAAGAT	TCTTTAGTTA	ATAGTTATTG	480
CTCTGGTGGC	TTGTGTGCAT	TTAGGAATC	TTACTCTGTT	TTTTAATGGA	GAAACGAAAC	540
CTACCTTTTT	TTCTGTGTTT	CCTTTTATGA	TATCACCTGC	TTGGAGGCGT	TTAGACTTTA	600
TCCACCTAAA	CTATTTCATG	TTACCAGACA	AGCTATAAGT	TTTATCCCCC	CCCCCGCGG	660
ACCTGNGGAC	AAAAGAAGCG	CTGATGAACT	GATTTAATCC	GTGTTTTTAT	ATATTACACA	720

TTGATGCTTC ATGGAGCTAA TATCTTTGGT TAAATTTTCAT GTATATATAT ACCCTTCCCT	780
CTTGTGATGG CAGTGGCCCC TCGTTTAATT AGCGTACTTA ATTATCTGAT GGATACTGTA	840
TGCTTGGCAG ATGATGTCAT CAGATTATAC CATTGTGTGT GCTCTACAAA ATAAAAAACC	900
TCTATTTATG TTCATCTTTT TGGTAACAAG TAACTAATTG ATGCGCTATG TTGACAGGCG	960
ATGCATTACA CAACTTACGA ACTAGCTTGC AAGATCCCAA CAATGTCTTG CAGAGCTGGG	1020
ATCCAACCCCT TGTGAACCCCT TGCACATGGT TTCATGTGAC ATGTAACAAT GAAAACAGTG	1080
TTATAAGAGT GTAGGTCAC TCCCTTATTA ATTTTTTTAG CAAGTTACGA ATATTTACTC	1140
AATTGAGCAG ATGTCCTCTT AAATATTTTT CTTTAATTTT TTAGCTAAGC GGAGCATCTA	1200
TCTTAAGTAT CTCTACTGAA TTTAAGACAT AATACATTTT TTTAAAAAAT CTATTAGAGT	1260
GTTTTTTCCG CACAGCGCAC ATATATCTTT TTTCTGGTAA TTCAGACAAC CTTTCTCCCG	1320
ACGATAAAAT AATATAAGAT TAACTCCTTG AACTAATTTT TTATTTTTCT TTTCTTTTTA	1380
TGTTCTTTGC AGAAAGTTTC TTATGGTCTT TTGTGAAAAG TACATTCTAT GATAATTTTT	1440
TGGCAACTCA TATAAATTTA TATATATTCC ATGTAGTTAT AAGTTAAAAA AAGCTTCCTA	1500
TTAATTCCAA GATAGAGGTT CATTTTTATA GTTTGGGCAT CCATGAGTTT TTGAAAATGT	1560
CAGAAATTTT GTTGAGTTAA TTTTACTTAC CAACTTTTAT GCGGTCAATG AGTGATCTTG	1620
GGAATGCAGC ATTATCTGGT CAATTGGTTC CTCTTGGCCA GTTGAAAAAT TTACAATACT	1680
TGTAAGACCA TATCACTTGG AATGCTTTAG TTTTTATACA GCACAATGCT TTCAATATCT	1740
GTTAAAAGTG TGAAAAAGTT GACTTTCTAG CTTCAGCAGT TGTTCGGATA ATATCTATGA	1800

00490799-1599

AGCACTTAAA AGGCTGGGCA ATTTTITGT TATTATTICA AATATTGTTA ATTGTTACTA 1860
CTTAATATGA TAAACTGATT TAACTCCTCA TGATTGGTCT CAGTCCAATG TGCCCTCATT 1920
AGTCACATNA TAAAAITGGEN GGGTTGGACA AATATAACTT CTTTICTTAA GGTCCAGAAA 1980
GAGCACTTAT CAACCTTGTC TAGCGCATAA CGTCACAGTG GGTGAGTCAC GGGCTATCCA 2040
GTTTGGGGAG GTTTTAATGA GCACTTATTT ACCTTGCTCT TTAACGCTCT GAGGATGTTA 2100
TTAAAGTCTG CATCAITTCAG AGTTTAAATT AGCACTTICA GTTGTATTAT GAATGGTACA 2160
TGAAAGATAC ATATCTTAAT GTTCCTATGC CTGTTTCAAC ATGCTCTTAA TATTCTGTTA 2220
TCTTTGTCAT CTTAAAAATG GCACTGATTA AAATGTGAGA AAGGTAGTCT TCCAATACCA 2280
TTTCATGTAT ACCAGAGAAT ATCATAATTT TTTTAAATCA TAAGTTGGGC CCTAGAGTTT 2340
TCTCAGTATT GGTCTATTTA TATTTTCCAC CATTTAGAAC TGTTGTGTC GATGAAAATC 2400
TTGGACTTCC ACAGAAGATC TTATAGTAAA AGTATTCTTT AGATCTGATG ATGAAAGTTG 2460
TCATGGTGTG GCCTGTCCCA GAATTTAAAT CAATCCCATG TCACATGTTT GTTGATCTGA 2520
CTACTCACTG TTAATCGAAG AGTAACTATT TGTGAATTAA ATGCTTTTTT TTTTGTCTTT 2580
CATGCTTAGC GTTATAAAGG TCTACGCTCG ACTATGGTTT TTAACATGTT ATAGTTTTGT 2640
ACTGACAAGT TTAAAGTTTC TCTTGTTTAC GAATTAAGAA TATATAATAT AAAACGCTTT 2700
AACTTTCTCT GTGGAAGGTG TTCTTACCTT TTTATATATA TATATAGATA CTCAGACTCT 2760
GCTGGCAATT ATATCTTACG AACTTACGAG TATACAGAAC TTGTATATTA GGTTGAGATG 2820
AGTGGCTGTA GTAGAACACC TTAAGCAAGA ACTTAATCAT GAGGTTTCAA CCTTTTAACT 2880
TTCTTTTTAG ATTTTTTCAA GTTTATGGAA AATTGTACCT CATGATCGTG GTTCTTTTCC 2940

00100700 11500
35511 36200160

ATAAACTTTC CATATAAGTC CGTTTCTTGA CGTTTTTCATG TAAGCTGTTG ACGAGTGAAT 3000

ATTAGCGGTT CTTTCAATAA TCATAATGTG TCTCACTTTG ATGAGGCCTG TACTTATTAT 3060

TGCACCTTGC ACTTAACCTT GATCCTCATG TCATCTTGAT TGTCAATAGT TACTAACCGA 3120

GTTGAACATG GTTTATCATG TCTTTTGAGG TAACAATGTA GCTTTCACCT CTGTCCCTGA 3180

TATAGGTTTA AGGCTTGCAC CTCCCACTAG CCTTTCGTTG TTTTATTAC AGTTCACACA 3240

CCTACTAGCA CTGTTACCT CTAGTCTTTT GTCGCAAAT AGTAAGAAGT TTCTTTGCA 3300

TAATAGTGA TGATCATTTA AGAAATAGTG AATCAAATTA TCGTGTATT GTGTTGTAC 3360

TTTGGAATTA AATGAGTTGC TGAACATTGT TGCTGTTTAT CGTTGTCAAG GCTTTGCCAA 3420

GGAAGGCGAT TAGTAAGAGT GGGCATCCAA GCGCCTTTAT CTTGAAGGGG CGGGCGGCAC 3480

GTGTGGATT CTGGGTGTCT ATTAGAGGAC ATTATCTATA TATACTGATT ATTTATTAGA 3540

ATATAAATCA ACTACTATAT TTTTCTTTGT AATGTTTATA TAGAAATCCC ACTCGTAAAC 3600

TTGACAAATA CCATTGAAAT ATTTGAACCT AATTAATTAG TAGTGTCAGG TTTAAATTCA 3660

AACTCATTTA ATTTTACTTT AAAAAATAAT TCTATATGAA TCGTAACAGT ATAAATATAT 3720

TAAATTACAT GTATGTGTGC CTATATATAG CTGAATGTCT AATAGACTCC AAGACGGCTG 3780

CTCTTACTGC CTAGGCGTCC AGGCAGTTCA CTGATGCTTA CCTTGACAAA TATGGGGTTC 3840

GTATGACATT GTTGGGGATC CCTATCACTG GATTCCTGTT TTGCTGACCC TCTGTTCAAT 3900

TGATTTTCAT TGATGTAGTA TTAAGTATT TATAAATATT CTTTATTGCA ATAATTAAAC 3960

TGGAGTTTAA CAATGACAGG GAGCTTTACA GCAATAACAT AAGTGGACCA ATTCTAGTG 4020

00400700-44000

ATCTTGGGAA TCTGACAAAT TTGGTGAGCT TGGACCTATA CATGAATAGC TTCTCTGGAC	4080
CTATACCGGA CACATTAGGA AAGCTTACAA GGCTAAGATT CTTGTATGAC TACAAATCTT	4140
CACTAGTTTT TAACTTAATG CAATTTGATT ATCCTTTCAA GTGATTGATT ATATCACAAA	4200
TTACTGGATA GCGTCTCAA CAACAACCTGC CTCTCTGGTC CAATCCAAT GTCACTGACT	4260
AATATTACAA CTCTTCAAGT CCTGTAAGTA TTCCGACCTT TCCAGATAGT TTTGTTGTG	4320
TGGATGTTTC AATTTTAATA CTAAATATGT TCATCAGGGA TTTATCAAAC AATCGGCTAT	4380
CAGGACCAGT ACCGGATAAT GGCTCATTTT CTTTGTITAC ACCTATCAGG TTTAATGCTA	4440
GTAATATCTT TAATATTATG GTTCTTACTT CTA CTGCGAA AGCTATGATA ATATTTTTTT	4500
TTCTCTTCAT ATATTATCAC TTTCGCAGTT TTGGCAATAA TTTGAATTTA TGTGGACCTG	4560
TAACTGGGAG GCCCTGCCCT GGATCTCCCC CATTTTCTCC ACCACCTCCG TTCATCCCAC	4620
CATCAACAGT ACAGCCTCCA GGTGATTAG TTTTATATT AATCCCGTA ATTAATTTTA	4680
TGACTGTAAA AATTGGTGTT AATTTACCA GTTGC GAATA AAGTATTTTC CTTCTTTCTC	4740
TTCTTATTAT TATGAAGGAC AAAATGGTCC CACTGGAGCT ATTGCTGGGG GAGTAGCTGC	4800
TGGTGCTGCT TTA CTGTTG CTGCACCTGC AATGGCATT GCATGGTGGC GGAGAAGAAA	4860
ACCGCGAGAA CATTTCTTTG ATGTGCCAGG TTAGTCTGT AAATAGATAT CTATTGAAGC	4920
GCTTACTGTC TGTGGACTTT GTTTCACTG TCATTAGTAA ACTTCAGCTG AAGAGGACCC	4980
AGAAGTGCAC CTTGGTCAAC TGAAGAGGTT TTCTCTGCGA GAATTGCAAG TCGCAACGGA	5040
TACTTTTAGT ACCATCCTTG GAAGAGGTGG ATTTGGTAAG GTGTATAAGG GACGCCTTGC	5100
TGATGGCTCA CTTGTAGCAG TTAAGGCT TAAAGAAGAA CGAACACCAG GTGGCGAGCT	5160

00190799-4499
009777-00208100

GCAGTTTCAA ACAGAAGTGG AAATGATTAG CATGGCTGTG CATCGAAATC TTCTGCGTCT	5220
ACGTGGTTTC TGCATGACAC CTACCGAGCG GCTTCTTGTA TATCCATACA TGGCTAATGG	5280
AAGTGTGCG TCATGTTTAA GAGGTATCTC AGTTACAATT ACCATAACTT GCCAGAAGTT	5340
TGTTTGATTA AAAATGAAAT ATAACTCCCT ACACTATGTT AAGGTGTTAT AATTTCTGAG	5400
CAGATCTTAT TTCCCATTC AAGATAACCAG TTATTATTGT TTTTCTGTG ATTGATACCG	5460
GTTATATTTT TTTCTTGAT TGGTTATAT GCAAGGATTT CGAGTCTAAT AAGTTATCAA	5520
ACTGGATGCT ATGTTTATTC TGCAATTGAA TTCTTGCTTC ATGTGCCAAA ATATATATGA	5580
TTCAACTTGG AATCATCTTA TAATATACTG TGTAAGTCA GCTGTGACT TTCATCATT	5640
ATTAGTCTTC ATAAATCAGA ATCTGCCTAG TGAGCTTTAC CGACATACTC TAAACCTTTC	5700
TTATGGCCCT GTATATAATC GTCCCACTTA CTTTATTCAG TTTGTCTGCT CTCIGAATTT	5760
TTGATCTGTA CATTGTGATG TCTTGTTTTT ATCAAATGTA GAGCGTCAGC CATCAGAACC	5820
TCCCCCTGAT TGGCCAATA GGGAGAGGAT TGCACTAGGA TCTTCTAGGG GCGTATCTAA	5880
ATTGCATGAC CATTGTGATC CCAAGATTAT CCATCGCGAT GTAAAAGCTG CAAATATATT	5940
ATTGGACGAA GAATTTGAGG CTGTTGTAGG TGATTTTGGG TTAGCTAGGC TCATGGATTA	6000
CAAGGATACC CATGTTACGA CTGCTGTAAG GGGTACCATT GGGCACATAG CTCCCGAGTA	6060
CCTCTCGACT GGAAAGTCAT CAGAGAAGAC CGATGTCCTT GGTATGTTG TAATGCTCCT	6120
AGAGCTCATT ACTGGACAGA GGGCTTTTGA TCTTGCTCGC CTGCGAACG ATGATGATGT	6180
TATGTTGTTG GATTGGGTAT GTGTCCCGGG TGTTCTTTG GTTAATTATT TCACATATTA	6240

GTGCTTACTA CTTTGTGTG GCCCTTGT TTTATTTCT GCTGTATTT GATTCTTAGT 6300
CATGTTATGC ATATTGACCT GCTTTGCAAT GTCTTTTAGG TTAAAAGCCT TTTGAAAGAG 6360
AAAAAGTTGG AGATGCTGGT CGATCCTGAC CTGCAGAACA ATTACATTGA CACAGAAGTT 6420
GAGCAGCTTA TTCAAGTAGC ATTACTCTGT ACCCAGGGTT CGCCAATGGA GCGGCCTAAG 6480
ATGTCAGAGG TAGTCCGAAT GCTTGAAGGT GATGGCCTTG CAGAAAAGTG GGACGAGTGG 6540
CAAAAAGTTG AAGTCATCCA TCAAGACGTA GAATTAGCTC CACATCGAAC TTCTGAATGG 6600
ATCCTAGACT CGACAGATAA CTTCGATGCT TTGAATTAT CTGGTCCAAG ATAAACAGCA 6660
TATAAAATGT AATGAAATTA ATATTTTTTA TGGTT 6695

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1815 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Daucus carota*

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 94..1752

09150700 14500

GACAAATACC	ATTGAAATAT	TTGAACCTAA	TTAATTAGTA	GTCACGGTT	TAAATTCAAA	60
CTCATTTAAT TTTACTTTAA AAAATAATTC TAT ATG AAT CGT AAC AGT ATA AAT						114
Met Asn Arg Asn Ser Ile Asn						
1 5						
ATA TTA AAT TAC ATG CAG TTC ACT GAT GCT TAC CTT GAC AAA TAT GGG						162
Ile Leu Asn Tyr Met Gln Phe Thr Asp Ala Tyr Leu Asp Lys Tyr Gly						
10 15 20						
GTT CTT ATG ACA TTG GAG CTT TAC AGC AAT AAC ATA AGT GGA CCA ATT						210
Val Leu Met Thr Leu Glu Leu Tyr Ser Asn Asn Ile Ser Gly Pro Ile						
25 30 35						
CCT AGT GAT CTT GGG AAT CTG ACA AAT TTG GTG AGC TTG GAC CTA TAC						258
Pro Ser Asp Leu Gly Asn Leu Thr Asn Leu Val Ser Leu Asp Leu Tyr						
40 45 50 55						
ATG AAT AGC TTC TCT GGA CCT ATA CCG GAC ACA TTA GGA AAG CTT ACA						306
Met Asn Ser Phe Ser Gly Pro Ile Pro Asp Thr Leu Gly Lys Leu Thr						
60 65 70						
AGG CTA AGA TTC TTG CGT CTC AAC AAC AAC AGC CTC TCT GGT CCA ATT						354
Arg Leu Arg Phe Leu Arg Leu Asn Asn Asn Ser Leu Ser Gly Pro Ile						
75 80 85						
CCA ATG TCA CTG ACT AAT ATT ACA ACT CTT CAA GTC CTG GAT TTA TCA						402
Pro Met Ser Leu Thr Asn Ile Thr Thr Leu Gln Val Leu Asp Leu Ser						
90 95 100						
AAC AAT CGG CTA TCA GGA CCA GTA CCG GAT AAT GGC TCA TTT TCT TTG						450
Asn Asn Arg Leu Ser Gly Pro Val Pro Asp Asn Gly Ser Phe Ser Leu						
105 110 115						
TTT ACA CCT ATC AGT TTT GCC AAT AAT TTG AAT TTA TGT GGA CCC GTA						498

[illegible]

265	270	275	
AGC ATG GCT GTG CAT CGA AAT CTT CTG CGT CTA CGT GGT TTC TGC ATG			978
Ser Met Ala Val His Arg Asn Leu Leu Arg Leu Arg Gly Phe Cys Met			
280	285	290	295
ACA CCA ACA GAG CGG CTT CTT GTA TAT CCA TAC ATG GCT AAT GGA AGT			1026
Thr Pro Thr Glu Arg Leu Leu Val Tyr Pro Tyr Met Ala Asn Gly Ser			
300	305		310
GTT GCG TCG TGT TTA AGA GAG CGT CAG CCA TCA GAA CCT CCC CTT GAT			1074
Val Ala Ser Cys Leu Arg Glu Arg Gln Pro Ser Glu Pro Pro Leu Asp			
315	320		325
TGG CCA ACT AGG AAG AGG ATT GCA CTA GGA TCT GCT AGG GGG CTT TCT			1122
Trp Pro Thr Arg Lys Arg Ile Ala Leu Gly Ser Ala Arg Gly Leu Ser			
330	335		340
TAT TTG CAT GAC CAT TGT GAT CCC AAG ATT ATC CAT CGT GAT GTA AAA			1170
Tyr Leu His Asp His Cys Asp Pro Lys Ile Ile His Arg Asp Val Lys			
345	350		355
GCT GCA AAT ATA TTA TTG GAC GAA GAA TTT GAG GCT GTT GTA GGT GAT			1218
Ala Ala Asn Ile Leu Leu Asp Glu Glu Phe Glu Ala Val Val Gly Asp			
360	365	370	375
TTT GGG TTA GCT AGG CTC ATG GAT TAC AAG GAT ACC CAT GTT ACA ACT			1266
Phe Gly Leu Ala Arg Leu Met Asp Tyr Lys Asp Thr His Val Thr Thr			
380	385		390
GCT GTA AGG GGT ACC TTG GGC TAC ATA GCT CCC GAG TAC CTC TCG ACT			1314
Ala Val Arg Gly Thr Leu Gly Tyr Ile Ala Pro Glu Tyr Leu Ser Thr			
395	400		405
GGA AAG TCA TCA GAG AAG ACC GAT GTC TTT GGT TAT GGG ATT ATG CTC			1362
Gly Lys Ser Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly Ile Met Leu			
410	415		420

00100700-11500

TTA GAG CTC ATT ACT GGA CAG AGA GCT TTT GAT CTT GCT CGC CTT GCG.	1410
Leu Glu Leu Ile Thr Gly Gln Arg Ala Phe Asp Leu Ala Arg Leu Ala	
425 430 435	
AAC GAT GAT GAT GTT ATG TTG TTG GAT TGG GTT AAA AGC CTT TTG AAA	1458
Asn Asp Asp Asp Val Met Leu Leu Asp Trp Val Lys Ser Leu Leu Lys	
440 445 450 455	
GAG AAA AAG TTG GAG ATG CTG GTC GAT CCT GAC CTG GAG AAC AAT TAC	1506
Glu Lys Lys Leu Glu Met Leu Val Asp Pro Asp Leu Glu Asn Asn Tyr	
460 465 470	
ATT GAC ACA GAA GTT GAG CAG CTT ATT CAA GTA GCA TTA CTC TGT ACC	1554
Ile Asp Thr Glu Val Glu Gln Leu Ile Gln Val Ala Leu Leu Cys Thr	
475 480 485	
CAG GGT TCG CCA ATG GAG CGG CCT AAG ATG TCA GAG GTA GTC CGA ATG	1602
Gln Gly Ser Pro Met Glu Arg Pro Lys Met Ser Glu Val Val Arg Met	
490 495 500	
CTT GAA GGT GAT GGC CTT GCA GAA AAG TGG GAC GAG TGG CAA AAA GTA	1650
Leu Glu Gly Asp Gly Leu Ala Glu Lys Trp Asp Glu Trp Gln Lys Val	
505 510 515	
GAA GTC ATC CAT CAA GAC GTA GAA TTA GCT CCA CAT CGA ACT TCT GAA	1698
Glu Val Ile His Gln Asp Val Glu Leu Ala Pro His Arg Thr Ser Glu	
520 525 530 535	
TGG ATC CTA GAC TCG ACA GAT AAC TTG CAT GCT TTT GAA TTA TCT GGT	1746
Trp Ile Leu Asp Ser Thr Asp Asn Leu His Ala Phe Glu Leu Ser Gly	
540 545 550	
CCA AGA TAAACAGCAT ATAAATGTG AATGAAATTA ATATTTTTTA TGGTTAAAAA	1802
Pro Arg	

AAAAAAAAAA AAA

1815

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 553 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Asn Arg Asn Ser Ile Asn Ile Leu Asn Tyr Met Gln Phe Thr Asp
1 5 10 15

Ala Tyr Leu Asp Lys Tyr Gly Val Leu Met Thr Leu Glu Leu Tyr Ser
20 25 30

Asn Asn Ile Ser Gly Pro Ile Pro Ser Asp Leu Gly Asn Leu Thr Asn
35 40 45

Leu Val Ser Leu Asp Leu Tyr Met Asn Ser Phe Ser Gly Pro Ile Pro
50 55 60

Asp Thr Leu Gly Lys Leu Thr Arg Leu Arg Phe Leu Arg Leu Asn Asn
65 70 75 80

Asn Ser Leu Ser Gly Pro Ile Pro Met Ser Leu Thr Asn Ile Thr Thr
85 90 95

Leu Gln Val Leu Asp Leu Ser Asn Asn Arg Leu Ser Gly Pro Val Pro
100 105 110

Asp Asn Gly Ser Phe Ser Leu Phe Thr Pro Ile Ser Phe Ala Asn Asn
115 120 125

00100709-14600

Leu	Asn	Leu	Cys	Gly	Pro	Val	Thr	Gly	Arg	Pro	Cys	Pro	Gly	Ser	Pro
130						135						140			
Pro	Phe	Ser	Pro	Pro	Pro	Pro	Phe	Ile	Pro	Pro	Ser	Thr	Val	Gln	Pro
145					150					155					160
Pro	Gly	Gln	Asn	Gly	Pro	Thr	Gly	Ala	Ile	Ala	Gly	Gly	Val	Ala	Ala
				165					170					175	
Gly	Ala	Ala	Leu	Leu	Phe	Ala	Ala	Pro	Ala	Met	Ala	Phe	Ala	Trp	Trp
			180					185					190		
Arg	Arg	Arg	Lys	Pro	Arg	Glu	His	Phe	Phe	Asp	Val	Pro	Ala	Glu	Glu
		195					200						205		
Asp	Pro	Glu	Val	His	Leu	Gly	Gln	Leu	Lys	Arg	Phe	Ser	Leu	Arg	Glu
	210					215					220				
Leu	Gln	Val	Ala	Thr	Asp	Thr	Phe	Ser	Thr	Ile	Leu	Gly	Arg	Gly	Gly
225					230					235					240
Phe	Gly	Lys	Val	Tyr	Lys	Gly	Arg	Leu	Ala	Asp	Gly	Ser	Leu	Val	Ala
				245					250					255	
Val	Lys	Arg	Leu	Lys	Glu	Glu	Arg	Thr	Pro	Gly	Gly	Glu	Leu	Gln	Phe
		260						265					270		
Gln	Thr	Glu	Val	Glu	Met	Ile	Ser	Met	Ala	Val	His	Arg	Asn	Leu	Leu
		275					280					285			
Arg	Leu	Arg	Gly	Phe	Cys	Met	Thr	Pro	Thr	Glu	Arg	Leu	Leu	Val	Tyr
	290					295					300				
Pro	Tyr	Met	Ala	Asn	Gly	Ser	Val	Ala	Ser	Cys	Leu	Arg	Glu	Arg	Gln
305					310					315					320
Pro	Ser	Glu	Pro	Pro	Leu	Asp	Trp	Pro	Thr	Arg	Lys	Arg	Ile	Ala	Leu

335

Trp Asp Glu Trp Gln Lys Val Glu Val Ile His Gln Asp Val Glu Leu
515 520 525

Ala Pro His Arg Thr Ser Glu Trp Ile Leu Asp Ser Thr Asp Asn Leu
530 535 540

His Ala Phe Glu Leu Ser Gly Pro Arg
545 550

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TTTTTTTTTT TGC

13

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

00180700-11500

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

10

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) **LENGTH:** 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

10

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TCAGCACAGG

10

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTTTTTTTTT TCTG

14

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

09190708-11600
SECRET

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTTTTTTTTT TCA

13

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GACATCGTCC

10

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

09160709-11500

(iii) ANTI-SENSE: NO

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

10

(2) INFORMATION FOR SEQ ID NO: 12:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

10

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) ANTI-SENSE: NO

(A) ORGANISM: primer

GGTGACTGTC

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TCTTGGACCA GATAATTC

18

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CTCTGATGAC TTTCAGTC

19

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AATGGCATTTC GCATGG

16

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

00180708:44500

(D) TOPOLOGY: unknown

(iii) **HYPOTHETICAL: NO**

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Daucus carota*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Ser Pro Pro Pro Pro

1

5

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Daucus carota*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

His Arg Asp Val Lys Ala Ala Asn

5

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Daucus carota*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Gly Thr Leu Gly Tyr Ile Ala Pro Glu

5

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4081 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(vii) IMMEDIATE SOURCE:

(ix) **FEATURE:**

(B) LOCATION: 1280..1367

(A) NAME/KEY: exon

(B) LOCATION: 1796..1928

(A) NAME/KEY: exon

(B) LOCATION: 2014..2085

(A) NAME/KEY: exon

(B) LOCATION: 2203..2346

(A) NAME/KEY: exon

(B) LOCATION: 2450..2521

(A) NAME/KEY: exon

(B) LOCATION: 2617..2688

(A) NAME/KEY: exon

(B) LOCATION: 2772..2884

(A) NAME/KEY: exon

(B) LOCATION: 3015..3146

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 3305..3646

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 3760..4081

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TCTAGAAACC TTTTGATCAT AATGAAAATA AAGAGTCCAT CCACCACATG GGGTAAGCAT	60
AATGTGTGAT ATTTAAAGGG TAACAAATGT AATCTGCTTT TTATTTTACT TTTTACCTCT	120
ACTCAAATG TATGGGCAGT TTTTTTTTTT TTTTAAATGA TAAGACAAGT ATCTGTTTAA	180
TGGTATTGTG ATGAAACAGT AGTAAAGTCA TATCGGGCAC GCCATACTAC TTCCACAGTG	240
GAACTTGGCC AAATTTTGTG TTTGCCGTCT CTACAGTTTC TTCCACCAAA TTTTTTGTG	300
ACAAACTCA AATCTTTCAA TCTCATCTCT GCCAAAGTTG GGTTTAGAAA GAATATCAGC	360
AAACACTAAT ATCTTTATTG TTGCATGGTT TATCAATCAC AAAATTCACA ACCATTGTAA	420
AAAAAAATTC ACATTTTGG TATGAGATTG CTCACATGAT AGTGAACCTC TTTAACATTT	480
TAACTTTACT TTCATAAATA CGGGATTACG AATCTTACTT GCATTAAAAA TTTAGAAAAG	540
GTTTTTCTAC TTAAAGAAAA AAGGGACCCA ACAGAGAGAG GTTTGACCAG GAGAAACGGG	600
TGCATAGCCT TAAGAGCTTT CAACTACTTT ACCCCAAACC CAAAGCGATG TCACTTTCAA	660
CCATCTCTTC TCTCCCCGA ACCCGTTTTT TTGACCGGTC AGTTGGGGCA GCAGCACCGT	720
TACGGGCAGC TTATATTCTT CGTCTTCTC CTCTACACCA CTGCATGCC ATAAATAAAG	780

00977-66203760

CCCGTTGAGA TCTTTAAAAA TATTAAATAA TATATCAACG AAAAAGCTAT TTTATTTCATA	840
AGAAGAAAAA GAGAGGAACA ACAACAACAC ACTAATCATA GTTCTCTGG CAGGCTTGTT	900
GTGCGGCTT AATAAAAAGC TCTTTTGTTA TTATTACTTC ACGTAGATTT TCCCCAAAAA	960
GCTCTTATTT TTTTGTTTAA AAAAAAAGT TTCATCTTTA TTCAACTTTT GTTTTACAGT	1020
GTGTGTGTGA GAGAGAGAGT GTGGTTTGAT TGAGGAAAGA CGACGACGAG AACGCCGGAG	1080
AATTAGGATT TTTATTTTAT TTTTACTCT TTGTTTGTTT TAATGCTAAT GGGTTTTTAA	1140
AAGGGTTATC GAAAAAATGA GTGAGTTTGT GTTGAGGTTG TCTCTGTAAA GTGTTAATGG	1200
TGGTGATTTT CGGAAGTTAG GGTCTCTCG GATCTGAAGA GATCAAATCA AGATTGAAA	1260
TTTAGCATTG TTGTTTGAAA TGGAGTCGAG TTATGTGGTG TTTATCTTAC TTTCACTGAT	1320
CTTACTTCOG AATCATTCAC TGTGGCTTGC TTCTGCTAAT TTGGAAGGTT CGTGGTTACT	1380
CAATTACTCA GCTTACTCG TTCTCAATT ACTTCTCGA TTCTTTTTTA TTGGAGGTG	1440
AATCGCTATC TTTAGTGTCT GCATTTTGAT TTATGAAAAT TGTGTGTGT CTTGTATTT	1500
GTAAGATTTA GTGGCTAGTA CTTTGAATAC ACTGTTTTGC TTTCTTGT CAGATCAACT	1560
TTGTATATTG TAAAGGCATG TTCTTTGGGT TGAAAAGCTG GGTATTTGA TATCTTAAGA	1620
TTGATGTTGT TGATCCAAAC ATTCTCTGAA AGACTTCATT TGTTTTTGGT TTTGTAAAGA	1680
ATTTGTTTAA TTATTAGCCT CTAATCTCAG AGAGGCTGT TTGAATAGTT CTCTCTTGAA	1740
ATTAGACTTT TCACCAATTG ATGCTAATTG TGTAGATTTG TTGTTCTTGT TATAGGTGAT	1800
GCTTTGCATA CTTTGAGGT TACTCTAGTT GATCCAAACA ATGTCTTGCA GAGCTGGGAT	1860
CCTACGCTAG TGAATCCTTG CACATGGTTC CATGTCACTT GCAACAACGA GAACAGTGTC	1920

09150799 11500

ATAAGAGTGT AAAGCTTTCT TCTACTAATC CCACMTTTTA AACMTTGACC TCAGCGTGGT	1980
TACCGACATT TTTGTTTCTT TTGTCAAATA CAGTGATTIG GGGAAATGCAG AGTTATCTGG	2040
CCATTTAGTT CCAGAGCTTG GTGTGCTCAA GAATTTGCAG TATTTGTAAG TTCCACTTAT	2100
GCATCATGCT TTAACAAAAC AAATCCAAGA TTTGACAGAA GAAGCACTGG AGTTACCTTT	2160
TGTAATTGAA ATCTTTTTAA CAAGTTTCTT ATTTTCTTAC AGGGAGCTTT ACAGTAACAA	2220
CATAACTGGC CCGATTCTTA GTAATCTTGG AAATCTGACA AACTTAGTGA GTTTGGATCT	2280
TTACTTAAAC AGCTTCTCCG GTCCTATTCC GGAATCATTG GGAAAGCTTT CAAAGCTGAG	2340
ATTTCTGTGA GTATACATAT GCTTTACCGG CTCAGTTACA GTCTTTGTTT AATCTTAGGT	2400
TTTGTTCCAA TTTTGTACTC TTTGCTGAAA ATTTTACATG CAAGAATAGC CGGCTTAACA	2460
ACAACAGTCT CACTGGGTCA ATTCTATGT CACTGACCAA TATTACTACC CTTCAAGTGT	2520
TGTGAGTCCT CTCATTAACT TTCATTTATG TCTACTTCAT TCTCCCTCAG TTGATTTGTT	2580
GAGTTAATGC ACTTAACCTT GATGGATGCA ACACAGAGAT CTATCAAATA ACAGACTCTC	2640
TGGTTCAGTT CCTGACAATG GCTCCTTCTC ACTCTTCACA CCCATCAGGT TCTATGATTT	2700
ATCCTCTTCA GTTATTTTAC TTGTTGTGTC AGTGTCTGAA CTTATTCTGA AACTTTTATT	2760
TCCTTGTGCA GTTTTGCTAA TAACTTAGAC CTATGTGGAC CTGTTACAAG TCACCCATGT	2820
CCTGGATCTC CCCCCTTTTC TCTCCACCA CCTTTTATTC AACCTCCCCC AGTTTCCACC	2880
CCGAGTAAGC CTCCTCTTTT TAGTTTACAT TATAGGAAAC AGAAGATGAA ATCTTTGCTT	2940
CTCTGTCAAT CCTTTTTCTC ATATAACTCA TCTTGCCAAT AAGGCAATAA CCAAATGATC	3000

00190799-14599

TAATTTGATT TCAGGTGGGT ATGGTATAAC TGGAGCAATA GCTGGTGGAG TTGCTGCAGG	3060
TGCTGCTTTG CTCTTTGCTG CTCTGCAAT AGCCTTTGCT TGGTGGCGAC GAAGAAAGCC	3120
ACTAGATATT TTCTTCGATG TGCTGGTGA GTTTATTATT CGCATTAGTT TCTGTCTTA	3180
GCCAGCAATT TTGTTTTGCA GAAAAGTATT GGAACAACCTG TTAATGAAAA TCAATACATA	3240
AGTCATTGTT TTTTAAGTTA CAAACTCTTT TGAGTAAAAT CTCGATTGCA AAATCTCTAT	3300
GCAGCCGAAG AAGATCCAGA AGTTCATCTG GGACAGCTCA AGAGGTTTTTC TTTGCGGGAG	3360
CTACAAGTGG CGAGTGATGG GTTTAGTAAC AAGAACATTT TGGGCAGAGG TGGGTTTGGG	3420
AAAGTCTACA AGGGACGCTT GGCAGACGGA ACTCTTGTTG CTGTCAAGAG ACTGAAGGAA	3480
GAGCGAACTC CAGGTGGAGA GCTCCAGTTT CAAACAGAAG TAGAGATGAT AAGTATGGCA	3540
GTTCATCGAA ACCTGTTGAG ATTACGAGGT TTCTGTATGA CACCGACCGA GAGATTGCTT	3600
GTGTATCCTT ACATGGCCAA TGGAAGTGTT GCTTCGTGTC TCAGAGGTAA AAATAAACA	3660
ATTAAACATC TTGTGCTCTC TCTCAATTAC TTTGACGTGA AGTGTTTTTT CATGTTTTCC	3720
TTTATGGGTT CATAATTGTT GGTTACACTA ATGACACAGA GAGGCCACCG TCACAACCTC	3780
CGCTTGATTG GCCAAGCGG AAGAGAATCG CGCTAGGCTC AGCTCGAGGT TTGTCTTACC	3840
TACATGATCA CTGCGATCCG AAGATCATTC ACCGTGACGT AAAAGCAGCA AACATCCTCT	3900
TAGACGAAGA ATTCGAAGCG GTTGTGGAG ATTTGGGGTT GGCAAAGCTA ATGGACTATA	3960
AAGACACTCA CGTGACAACA GCAGTCCGTG GCACCATCGG TCACATCGCT CCAGAATATC	4020
TCTCAACCGG AAAATCTTCA GAGAAAACCG ACGTTTTCGG ATACGGAATC ATGCTTCTAG	4080
A	4081

00100700-11500

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 494 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

Met Glu Ser Ser Tyr Val Val Phe Ile Leu Leu Ser Leu Ile Leu Leu
1 5 10 15

Pro Asn His Ser Leu Trp Leu Ala Ser Ala Asn Leu Glu Gly Asp Ala
 20 25 30

Leu His Thr Leu Arg Val Thr Leu Val Asp Pro Asn Asn Val Leu Gln
 35 40 45

Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val Thr
 50 55 60

Cys Asn Asn Glu Asn Ser Val Ile Arg Val Asp Leu Gly Asn Ala Glu
65 70 75 80

Leu Ser Gly His Leu Val Pro Glu Leu Gly Val Leu Lys Asn Leu Gln
 85 90 95

00190799-1150
009111-00208150

Glu Leu Tyr Ser Asn Asn Ile Thr Gly Pro Ile Pro Ser Asn Leu Gly
100 105 110

Asn Leu Thr Asn Leu Val Ser Leu Asp Leu Tyr Leu Asn Ser Phe Ser
115 120 125

Gly Pro Ile Pro Glu Ser Leu Gly Lys Leu Ser Lys Leu Arg Phe Leu
130 135 140

Arg Leu Asn Asn Asn Ser Leu Thr Gly Ser Ile Pro Met Ser Leu Thr
145 150 155 160

Asn Ile Thr Thr Leu Gln Val Leu Asp Leu Ser Asn Asn Arg Leu Ser
165 170 175

Gly Ser Val Pro Asp Asn Gly Ser Phe Ser Leu Phe Thr Pro Ile Ser
180 185 190

Phe Ala Asn Asn Leu Asp Leu Cys Gly Pro Val Thr Ser His Pro Cys
195 200 205

Pro Gly Ser Pro Pro Phe Ser Pro Pro Pro Phe Ile Gln Pro Pro
210 215 220

Pro Val Ser Thr Pro Ser Gly Tyr Gly Ile Thr Gly Ala Ile Ala Gly
225 230 235 240

Gly Val Ala Ala Gly Ala Ala Leu Leu Phe Ala Ala Pro Ala Ile Ala
245 250 255

Phe Ala Trp Trp Arg Arg Arg Lys Pro Leu Asp Ile Phe Phe Asp Val
260 265 270

Pro Ala Glu Glu Asp Pro Glu Val His Leu Gly Gln Leu Lys Arg Phe
275 280 285

Ser Leu Arg Glu Leu Gln Val Ala Ser Asp Gly Phe Ser Asn Lys Asn

00110700-11500

290	295	300
Ile Leu Gly Arg Gly Gly Phe Gly Lys Val Tyr Lys Gly Arg Leu Ala		
305	310	315 320
Asp Gly Thr Leu Val Ala Val Lys Arg Leu Lys Glu Glu Arg Thr Pro		
325	330	335
Gly Gly Glu Leu Gln Phe Gln Thr Glu Val Glu Met Ile Ser Met Ala		
340	345	350
Val His Arg Asn Leu Leu Arg Leu Arg Gly Phe Cys Met Thr Pro Thr		
355	360	365
Glu Arg Leu Leu Val Tyr Pro Tyr Met Ala Asn Gly Ser Val Ala Ser		
370	375	380
Cys Leu Arg Glu Arg Pro Pro Ser Gln Pro Pro Leu Asp Trp Pro Thr		
385	390	395 400
Arg Lys Arg Ile Ala Leu Gly Ser Ala Arg Gly Leu Ser Tyr Leu His		
405	410	415
Asp His Cys Asp Pro Lys Ile Ile His Arg Asp Val Lys Ala Ala Asn		
420	425	430
Ile Leu Leu Asp Glu Glu Phe Glu Ala Val Val Gly Asp Phe Gly Leu		
435	440	445
Ala Lys Leu Met Asp Tyr Lys Asp Thr His Val Thr Thr Ala Val Arg		
450	455	460
Gly Thr Ile Gly His Ile Ala Pro Glu Tyr Leu Ser Thr Gly Lys Ser		
465	470	475 480
Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly Ile Met Leu Leu		
485	490	

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1106 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 142..795

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TCGACCCACG CGTCGGTCCA ACTTCAATAA AGGGGAAACC AACGTAACCC TAATTTTGCT	60
TTCTCCTCTT TGTTCAGAAA ATTTTCCCTT TACTCTCAAA TTCCTTTTCG ATTTCCCTCT	120
CTTAAACCTC CGAAAGCTCA C ATG GCG TCT CGA AAC TAT CGG TGG GAG CTC	171
Met Ala Ser Arg Asn Tyr Arg Trp Glu Leu	
1 5 10	
TTC GCA GCT TCG TTA ACC CTA ACC TTA GCT TTG ATT CAC CTG GTC GAA	219
Phe Ala Ala Ser Leu Thr Leu Thr Leu Ala Leu Ile His Leu Val Glu	
15 20 25	
GCA AAC TCC GAA GGA GAT GCT CTC TAC GCT CTT CGC CGG AGT TTG ACA	267
Ala Asn Ser Glu Gly Asp Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr	
30 35 40	

00100700 111500

GAT CCA GAC CAT GTC CTC CAG AGC TGG GAT CCA ACT CTT GTT AAT CCT	315
Asp Pro Asp His Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro	
45 50 55	
TGT ACC TGG TTC CAT GTC ACC TGT AAC CAA GAC AAC CGC GTC ACT CGT	363
Cys Thr Trp Phe His Val Thr Cys Asn Gln Asp Asn Arg Val Thr Arg	
60 65 70	
GTG GAT TTG GGA AAT TCA AAC CTC TCT GGA CAT CTT GCG CCT GAG CTT	411
Val Asp Leu Gly Asn Ser Asn Leu Ser Gly His Leu Ala Pro Glu Leu	
75 80 85 90	
GGG AAG CTT GAA CAT TTA CAG TAT CTA GAG CTC TAC AAA AAC AAC ATC	459
Gly Lys Leu Glu His Leu Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile	
95 100 105	
CAA GGA ACT ATA CCT TCC GAA CTT GGA AAT CTG AAG AAT CTC ATC AGC	507
Gln Gly Thr Ile Pro Ser Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser	
110 115 120	
TTG GAT CTG TAC AAC AAC AAT CTT ACA GGG ATA GTT CCC ACT TTC TTG	555
Leu Asp Leu Tyr Asn Asn Asn Leu Thr Gly Ile Val Pro Thr Phe Leu	
125 130 135	
GGA AAA TTG AAG TCT CTG GTC TTT TTA CGG CTT AAT GAC AAC CGA TTG	603
Gly Lys Leu Lys Ser Leu Val Phe Leu Arg Leu Asn Asp Asn Arg Leu	
140 145 150	
ACC GGT CCA ATC CTA GAG CAC TCA CGG CAA TCC CAA GCC TTT AAA GTT	651
Thr Gly Pro Ile Leu Glu His Ser Arg Gln Ser Gln Ala Phe Lys Val	
155 160 165 170	
GTT GAC GTC TCA AGC AAT GAT TTG TGT GGG ACA ATC CCA ACA AAC GGA	699
Val Asp Val Ser Ser Asn Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly	
175 180 185	
CCC TTT GCT CAC ATT CCT TTA CAG AAC TTT GAG AAC AAC CCG AGA TTG	747

Pro Phe Ala His Ile Pro Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu
190 195 200

GAG GGA CCG GAA TTA CTC GGT CTT GCA AGC TAC GAC ACT AAC TGC ACC 795
Glu Gly Pro Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr
205 210 215

TGAAACAACCT GGCAAAACCT GAAAATGAAG AATTGGGGGG TGACCTTGTA AGAACAACCTC 855

ACCACTTTAT CAAATATCAC ATCTATTATG TAATAAGTAT ATATATGTAG TAAAAACAAA 915

AAAAATGAAG AATCGAATCG GTAATATCAT CTGGTCTCAA TTGAGAACTT CGAGGTCTGT 975

ATGTAAAATT TCTAAATGCG ATTTTCGCTT ACTGTAATGT TCGGTTGTGG GATTCTGAGA 1035

AGTAACATTT GTATTGGTAT GGTATCAAGT TGTTCGCTT TGTCTGCAA AAAAAAAAAA 1095

AAAAAAAAA A 1106

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 218 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

Met Ala Ser Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Thr
1 5 10 15
Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp
20 25 30

0019079-1190

Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu
35 40 45

Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val
50 55 60

Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser
65 70 75 80

Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu
85 90 95

Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser
100 105 110

Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn
115 120 125

Asn Leu Thr Gly Ile Val Pro Thr Phe Leu Gly Lys Leu Lys Ser Leu
130 135 140

Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Leu Glu
145 150 155 160

His Ser Arg Gln Ser Gln Ala Phe Lys Val Val Asp Val Ser Ser Asn
165 170 175

Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro
180 185 190

Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu
195 200 205

Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr
210 215

(2) INFORMATION FOR SEQ ID NO: 24:

00180790-1199
00000000-00000000

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 981 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 104..757

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AGTGTGAGTA ATTAGTTTG CTTTCTCCTC TTTGTTTCAGA AAATTTTCCC TTTACTCTCA	60
AATTCCTTTT CGATTTCCTT CTCTTAAACC TCCGAAAGCT CAC ATG GCG TCT CGA	115
Met Ala Ser Arg	
1	
AAC TAT CGG TGG GAG CTC TTC GCA GCT TCG TTA ACC CTA ACC TTA GCT	163
Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Thr Leu Thr Leu Ala	
5 10 15 20	
TTG ATT CAC CTG GTC GAA GCA AAC TCC GAA GGA GAT GCT CTC TAC GCT	211
Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp Ala Leu Tyr Ala	
25 30 35	
CTT CGC CGG AGT TTG ACA GAT CCA GAC CAT GTC CTC CAG AGC TGG GAT	259
Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu Gln Ser Trp Asp	
40 45 50	

001607090 145000

CCA ACT CTT GTT AAT CCT TGT ACC TGG TTC CAT GTC ACC TGT AAC CAA	307
Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val Thr Cys Asn Gln	
55 60 65	
GAC AAC CGC GTC ACT CGT GTG GAT TTG GGA AAT TCA AAC CTC TCT GGA	355
Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser Asn Leu Ser Gly	
70 75 80	
CAT CTT GCG CCT GAG CTT GGG AAG CTT GAA CAT TTA CAG TAT CTA GAG	403
His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu Gln Tyr Leu Glu	
85 90 95 100	
CTC TAC AAA AAC AAC ATC CAA GGA ACT ATA CCT TCC GAA CTT GGA AAT	451
Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser Glu Leu Gly Asn	
105 110 115	
CTG AAG AAT CTC ATC AGC TTG GAT CTG TAC AAC AAC AAT CTT ACA GGG	499
Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn Asn Leu Thr Gly	
120 125 130	
ATA GTT CCC ACT TCT TTG GGA AAA TTG AAG TCT CTG GTC TTT TTA CGG	547
Ile Val Pro Thr Ser Leu Gly Lys Leu Lys Ser Leu Val Phe Leu Arg	
135 140 145	
CTT AAT GAC AAC CGA TTG ACC GGT CCA ATC CCT AGA GCA CTC ACG GCA	595
Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg Ala Leu Thr Ala	
150 155 160	
ATC CCA AGC CTT AAA GTT GTT GAC GTC TCA AGC AAT GAT TTG TGT GGA	643
Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn Asp Leu Cys Gly	
165 170 175 180	
ACA ATC CCA ACA AAC GGA CCC TTT GCT CAC ATT CCT TTA CAG AAC TTT	691
Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro Leu Gln Asn Phe	
185 190 195	
GAG AAC AAC CCG AGA TTG GAG GGA CCG GAA TTA CTC GGT CTT GCA AGC	739

00944-00208160

50	55	60
Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser		
65	70	75 80
Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu		
85	90	95
Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser		
100	105	110
Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn		
115	120	125
Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu Lys Ser Leu		
130	135	140
Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg		
145	150	155 160
Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn		
165	170	175
Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro		
180	185	190
Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu		
195	200	205
Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr		
210	215	

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 789 base pairs

(B) TYPE: nucleic acid

00480769-11966

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..661

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

T CGA CCC ACG CGT CCG CGA AAC TAT CGG TGG GAG CTC TTC GCA GCT	46
Arg Pro Thr Arg Pro Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala	
1 5 10 15	
TCG TTA ATC CTA ACC TTA GCT TTG ATT CAC CTG GTC GAA GCA AAC TCC	94
Ser Leu Ile Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser	
20 25 30	
GAA GGA GAT GCT CTT TAC GCT CTT CGC CGG AGT TTA ACA GAT CCG GAC	142
Glu Gly Asp Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp	
35 40 45	
CAT GTT CTC CAG AGC TGG GAT CCA ACT CTT GTT AAT CCT TGT ACC TGG	190
His Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp	
50 55 60	
TTC CAT GTC ACC TGT AAC CAA GAC AAC CGC GTC ACT CGT GTG GAT TTG	238
Phe His Val Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu	
65 70 75	
GGG AAT TCA AAC CTC TCT GGA CAT CTT GCG CCT GAG CTT GGG AAG CTT	286
Gly Asn Ser Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu	

00100700-14500

80	85	90	95	
GAA CAT TTA CAG TAT CTA GAG CTC TAC AAA AAC AAC ATC CAA GGA ACT				334
Glu His Leu Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr				
100	105	110		
ATA CCT TCC GAA CTT GGA AAT CTG AAG AAT CTC ATC AGC TTG GAT CTG				382
Ile Pro Ser Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu				
115	120	125		
TAC AAC AAC AAT CTT ACA GGG ATA GTT CCC ACT TCT TTG GGA AAA TTG				430
Tyr Asn Asn Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu				
130	135	140		
AAG TCT CTG GTC TTT TTA CGG CTT AAT GAC AAC CGA TTG ACG GGG CCA				478
Lys Ser Leu Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro				
145	150	155		
ATC CCT AGA GCA CTC ACT GCA ATC CCA AGC CTT AAA GTT GTT GAT GTC				526
Ile Pro Arg Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val				
160	165	170	175	
TCA AGC AAT GAT TTG TGT GGA ACA ATC CCA ACA AAC GGA CCT TTT GCT				574
Ser Ser Asn Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala				
180	185	190		
CAC ATT CCT TTA CAG AAC TTT GAG AAC AAC CCG AGG TTG GAG GGA CCG				622
His Ile Pro Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro				
195	200	205		
GAA TTA CTC GGT CTT GCA AGC TAC GAC ACT AAC TGC ACC TGAAAAATT				671
Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr				
210	215	220		
GGCAAAACCT GAAAATGAAG AATTGGGGGG TGACCTTGTA AGAACACTTC ACCACTTTAT				731
CAAATATCAC ATCTACTATG TAATAAGTAT ATATATGTAG TCCAAAAAAA AAAAAAAA				789

00150700-11500

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 220 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Arg Pro Thr Arg Pro Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser
1 5 10 15
Leu Ile Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu
20 25 30
Gly Asp Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His
35 40 45
Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe
50 55 60
His Val Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly
65 70 75 80
Asn Ser Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu
85 90 95
His Leu Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile
100 105 110
Pro Ser Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr
115 120 125

00150709-11500

Asn Asn Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu Lys
130 135 140

Ser Leu Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile
145 150 155 160

Pro Arg Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser
165 170 175

Ser Asn Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His
180 185 190

Ile Pro Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu
195 200 205

Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr
210 215 220

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 894 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..675

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

00190709 141500

GGA CCG ATT CAA GCC TCC GAA GGG GAC GCT CTT CAC GCG CTT CGC CGG	48
Gly Pro Ile Gln Ala Ser Glu Gly Asp Ala Leu His Ala Leu Arg Arg	
1 5 10 15	
AGC TTA TCA GAT CCA GAC AAT GTT GTT CAG AGT TGG GAT CCA ACT CTT	96
Ser Leu Ser Asp Pro Asp Asn Val Val Gln Ser Trp Asp Pro Thr Leu	
20 25 30	
GTT AAT CCT TGT ACT TGG TTT CAT GTC ACT TGT AAT CAA CAC CAT CAA	144
Val Asn Pro Cys Thr Trp Phe His Val Thr Cys Asn Gln His His Gln	
35 40 45	
GTC ACT CGT CTG GAT TTG GGG AAT TCA AAC TTA TCT GGA CAT CTA GTA	192
Val Thr Arg Leu Asp Leu Gly Asn Ser Asn Leu Ser Gly His Leu Val	
50 55 60	
CCT GAA CTT GGG AAG CTT GAA CAT TTA CAA TAT CTG TAT GGA ATC ATC	240
Pro Glu Leu Gly Lys Leu Glu His Leu Gln Tyr Leu Tyr Gly Ile Ile	
65 70 75 80	
ACT CTT TTG CCT TTT GAT TAT CTG AAA ACA TTT ACA TTA TCA GTC ACA	288
Thr Leu Leu Pro Phe Asp Tyr Leu Lys Thr Phe Thr Leu Ser Val Thr	
85 90 95	
CAT ATA ACA TTT TGC TTT GAG TCA TAT AGT GAA CTC TAC AAA AAC GAG	336
His Ile Thr Phe Cys Phe Glu Ser Tyr Ser Glu Leu Tyr Lys Asn Glu	
100 105 110	
ATT CAA GGA ACT ATA CCT TCT GAG CTT GGA AAT CTG AAG AGT CTA ATC	384
Ile Gln Gly Thr Ile Pro Ser Glu Leu Gly Asn Leu Lys Ser Leu Ile	
115 120 125	
AGT TTG GAT CTG TAC AAC AAC AAT CTC ACC GGG AAA ATC CCA TCT TCT	432
Ser Leu Asp Leu Tyr Asn Asn Asn Leu Thr Gly Lys Ile Pro Ser Ser	
130 135 140	

00190799-11500

TTG GGA AAA TTG AAG TCA CTT GTT TTT TTG CGG CTT AAC GAA AAC CGA	480
Leu Gly Lys Leu Lys Ser Leu Val Phe Leu Arg Leu Asn Glu Asn Arg	
145 150 155 160	
TTG ACC GGT CCT ATT CCT AGA GAA CTC ACA GTT ATT TCA AGC CTT AAA	528
Leu Thr Gly Pro Ile Pro Arg Glu Leu Thr Val Ile Ser Ser Leu Lys	
165 170 175	
GTT GTT GAT GTC TCA GGG AAT GAT TTG TGT GGA ACA ATT CCA GTA GAA	576
Val Val Asp Val Ser Gly Asn Asp Leu Cys Gly Thr Ile Pro Val Glu	
180 185 190	
GGA CCT TTT GAA CAC ATT CCT ATG CAA AAC TTT GAG AAC AAC CTG AGA	624
Gly Pro Phe Glu His Ile Pro Met Gln Asn Phe Glu Asn Asn Leu Arg	
195 200 205	
TTG GAG GGA CCA GAA CTA CTA GGT CTT GCG AGC TAT GAC ACC AAT TGC	672
Leu Glu Gly Pro Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys	
210 215 220	
ACT TAAAAAGAAG TTGAAGAACC TATAAAGAAG AATGTTAGGT GACCTTGTA	725
Thr	
225	
GAACCTGTGA CCAAGTGTTC GTAAATCTAT ATAGAGCCTT GTTTCATGTT ATATATGAAA	785
GCTTTGAGAG ACAGTAACTT GCAATGTATT GGTATTGGTA GAAAAAGTTG AAATGAGAAT	845
TGCTTTGTAA TTGGATTGT GTTCTTATG TAACTTGAAT TTCTTATTA	894

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 225 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

00150799:44595

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Gly Pro Ile Gln Ala Ser Glu Gly Asp Ala Leu His Ala Leu Arg Arg
1 5 10 15

Ser Leu Ser Asp Pro Asp Asn Val Val Gln Ser Trp Asp Pro Thr Leu
20 25 30

Val Asn Pro Cys Thr Trp Phe His Val Thr Cys Asn Gln His His Gln
35 40 45

Val Thr Arg Leu Asp Leu Gly Asn Ser Asn Leu Ser Gly His Leu Val
50 55 60

Pro Glu Leu Gly Lys Leu Glu His Leu Gln Tyr Leu Tyr Gly Ile Ile
65 70 75 80

Thr Leu Leu Pro Phe Asp Tyr Leu Lys Thr Phe Thr Leu Ser Val Thr
85 90 95

His Ile Thr Phe Cys Phe Glu Ser Tyr Ser Glu Leu Tyr Lys Asn Glu
100 105 110

Ile Gln Gly Thr Ile Pro Ser Glu Leu Gly Asn Leu Lys Ser Leu Ile
115 120 125

Ser Leu Asp Leu Tyr Asn Asn Asn Leu Thr Gly Lys Ile Pro Ser Ser
130 135 140

Leu Gly Lys Leu Lys Ser Leu Val Phe Leu Arg Leu Asn Glu Asn Arg
145 150 155 160

Leu Thr Gly Pro Ile Pro Arg Glu Leu Thr Val Ile Ser Ser Leu Lys
165 170 175

00000000000000000000

Val Val Asp Val Ser Gly Asn Asp Leu Cys Gly Thr Ile Pro Val Glu
180 185 190

Gly Pro Phe Glu His Ile Pro Met Gln Asn Phe Glu Asn Asn Leu Arg
195 200 205

Leu Glu Gly Pro Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys
210 215 220

Thr
225

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1063 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 106..759

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

TOGACCCACG CGTCCGACGA AACCCCTAATT TTGCTTCCTC ATCTTGTTCA GAAAATTACT 60

CAAATTCTTA TTAGATTACT CTCTCTTCGA CCTCCGATAG CTCAC ATG GCG TCT 114
Met Ala Ser

09180799-11599

210

258

306

354

402

450

498

548

CGG CTT AAT GAC AAC CGA TTG ACG GGG CCA ATC CCT AGA GCA CTC ACT	594
Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg Ala Leu Thr	
150 155 160	
GCA ATC CCA AGC CTT AAA GTT GTT GAT GTC TCA AGC AAT GAT TTG TGT	642
Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn Asp Leu Cys	
165 170 175	
GGA ACA ATC CCA ACA AAC GGA CCT TTT GCT CAC ATT CCT TTA CAG AAC	690
Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro Leu Gln Asn	
180 185 190 195	
TTT GAG AAC AAC CCG AGG TTG GAG GGA CCG GAA TTA CTC GGT CTT GCA	738
Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu Gly Leu Ala	
200 205 210	
AGC TAC GAC ACT AAC TGC ACC TGAAAAAATT GGCAAAACCT GAAAATGAAG	789
Ser Tyr Asp Thr Asn Cys Thr	
215	
AATTGGGGGG TGACCTTGTA AGAACACTTC ACCACTTTAT CAAATATCAC ATCTACTATG	849
TAATAAGTAT ATATATGTAG TCCAAAAAAA AAATGAAGAA TCGAATCAGT AATATCATCT	909
GGTCTCAATT GAGAACTTTG AGGTCTGTGT ATGTAAAATT TCTAAATGCG ACTTTTCGCGT	969
ACTGTAATGT TCGGTGTGG GATTCTGAGA AGTAACATTT GTATTGGTAT GGTATCAAGT	1029
TGTTCTGCCT TGTCTGCAAA AAAAAAAAAA AAAA	1063

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 218 amino acids

(B) TYPE: amino acid

00160799-14500

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

Met Ala Ser Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Ile
1 5 10 15

Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp
20 25 30

Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu
35 40 45

Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val
50 55 60

Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser
65 70 75 80

Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu
85 90 95

Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser
100 105 110

Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn
115 120 125

Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu Lys Ser Leu
130 135 140

Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg
145 150 155 160

Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn

00100700-11600

(2) INFORMATION FOR SEQ ID NO: 32:

(A) LENGTH: 2089 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

(B) CLONE: SERK gene cDNA

(ix) **FEATURE:**

(A) NAME/KEY: CDS

(B) LOCATION: 195..2069

GGATTTTAT TTTATTTTTT ACTCTTTGTT TGTTTTAATG CTAATGGGTT TTTAAAAGGG	60
TTATCGAAAA AATGAGTGAG TTTGIGTGA GGTGTCCTCT GTAAAGTGTT AATGGTGGTG	120
ATTTTCGGAA GTTAGGGTTT TCTCGGATCT GAAGAGATCA AATCAAGATT CGAAATTTAC	180
CATTGTTGTT TGAA ATG GAG TCG AGT TAT GTG GTG TTT ATC TTA CTT TCA	230
Met Glu Ser Ser Tyr Val Val Phe Ile Leu Leu Ser	
1 5 10	
CTG ATC TTA CTT CCG AAT CAT TCA CTG TGG CTT GCT TCT GCT AAT TTG	278
Leu Ile Leu Leu Pro Asn His Ser Leu Trp Leu Ala Ser Ala Asn Leu	
15 20 25	
GAA GGT GAT GCT TTG CAT ACT TTG AGG GTT ACT CTA GTT GAT CCA AAC	326
Glu Gly Asp Ala Leu His Thr Leu Arg Val Thr Leu Val Asp Pro Asn	
30 35 40	
AAT GTC TTG CAG AGC TGG GAT CCT ACG CTA GTG AAT CCT TGC ACA TGG	374
Asn Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp	
45 50 55 60	
TTC CAT GTC ACT TGC AAC AAC GAG AAC AGT GTC ATA AGA GTT GAT TTG	422
Phe His Val Thr Cys Asn Asn Glu Asn Ser Val Ile Arg Val Asp Leu	
65 70 75	
GGG AAT GCA GAG TTA TCT GGC CAT TTA GTT CCA GAG CTT GGT GTG CTC	470
Gly Asn Ala Glu Leu Ser Gly His Leu Val Pro Glu Leu Gly Val Leu	
80 85 90	
AAG AAT TTG CAG TAT TTG GAG CTT TAC AGT AAC AAC ATA ACT GGC CCG	518
Lys Asn Leu Gln Tyr Leu Glu Leu Tyr Ser Asn Asn Ile Thr Gly Pro	
95 100 105	
ATT CCT AGT AAT CTT GGA AAT CTG ACA AAC TTA GTG AGT TTG GAT CTT	566

Ile	Pro	Ser	Asn	Leu	Gly	Asn	Leu	Thr	Asn	Leu	Val	Ser	Leu	Asp	Leu		
110						115					120						
TAC	TTA	AAC	AGC	TTC	TCC	GGT	CCT	ATT	CCG	GAA	TCA	TTG	GGA	AAG	CTT	614	
Tyr	Leu	Asn	Ser	Phe	Ser	Gly	Pro	Ile	Pro	Glu	Ser	Leu	Gly	Lys	Leu		
125					130					135					140		
TCA	AAG	CTG	AGA	TTT	CTC	CGG	CTT	AAC	AAC	AAC	AGT	CTC	ACT	GGG	TCA	662	
Ser	Lys	Leu	Arg	Phe	Leu	Arg	Leu	Asn	Asn	Asn	Ser	Leu	Thr	Gly	Ser		
				145				150						155			
ATT	CCT	ATG	TCA	CTG	ACC	AAT	ATT	ACT	ACC	CTT	CAA	GTG	TTA	GAT	CTA	710	
Ile	Pro	Met	Ser	Leu	Thr	Asn	Ile	Thr	Thr	Leu	Gln	Val	Leu	Asp	Leu		
				160				165					170				
TCA	AAT	AAC	AGA	CTC	TCT	GGT	TCA	GTT	CCT	GAC	AAT	GGC	TCC	TTC	TCA	758	
Ser	Asn	Asn	Arg	Leu	Ser	Gly	Ser	Val	Pro	Asp	Asn	Gly	Ser	Phe	Ser		
				175				180					185				
CTC	TTC	ACA	CCC	ATC	AGT	TTT	GCT	AAT	AAC	TTA	GAC	CTA	TGT	GGA	CCT	806	
Leu	Phe	Thr	Pro	Ile	Ser	Phe	Ala	Asn	Asn	Leu	Asp	Leu	Cys	Gly	Pro		
	190					195						200					
GTT	ACA	AGT	CAC	CCA	TGT	CCT	GGA	TCT	CCC	CCG	TTT	TCT	CCT	CCA	CCA	854	
Val	Thr	Ser	His	Pro	Cys	Pro	Gly	Ser	Pro	Pro	Phe	Ser	Pro	Pro	Pro		
205					210					215				220			
CCT	TTT	ATT	CAA	CCT	CCC	CCA	GTT	TCC	ACC	CCG	AGT	GGG	TAT	GGT	ATA	902	
Pro	Phe	Ile	Gln	Pro	Pro	Pro	Val	Ser	Thr	Pro	Ser	Gly	Tyr	Gly	Ile		
				225				230					235				
ACT	GGA	GCA	ATA	GCT	GGT	GGA	GTT	GCT	GCA	GGT	GCT	GCT	TIG	CCC	TTT	950	
Thr	Gly	Ala	Ile	Ala	Gly	Gly	Val	Ala	Ala	Gly	Ala	Ala	Leu	Pro	Phe		
				240				245					250				
GCT	GCT	CCT	GCA	ATA	GCC	TTT	GCT	TGG	TGG	CGA	CGA	AGA	AGC	CCA	CTA	998	
Ala	Ala	Pro	Ala	Ile	Ala	Phe	Ala	Trp	Trp	Arg	Arg	Arg	Ser	Pro	Leu		

00100300-11500

265

CCG CTT GAT TGG CCA ACG CGG AAG AGA ATC GCG CTA GGC TCA GCT CGA 1430
Pro Leu Asp Trp Pro Thr Arg Lys Arg Ile Ala Leu Gly Ser Ala Arg
400 405 410

GGT TTG TCT TAC CTA CAT GAT CAC TGC GAT CCG AAG ATC ATT CAC CGT	1478
Gly Leu Ser Tyr Leu His Asp His Cys Asp Pro Lys Ile Ile His Arg	
415 420 425	
GAC GTA AAA GCA GCA AAC ATC CTC TTA GAC GAA GAA TTC GAA GCG GTT	1526
Asp Val Lys Ala Ala Asn Ile Leu Leu Asp Glu Glu Phe Glu Ala Val	
430 435 440	
GTT GGA GAT TTC GGG TTG GCA AAG CTT ATG GAC TAT AAA GAC ACT CAC	1574
Val Gly Asp Phe Gly Leu Ala Lys Leu Met Asp Tyr Lys Asp Thr His	
445 450 455 460	
GTG ACA ACA GCA GTC CGT GGC ACC ATC GGT CAC ATC GCT CCA GAA TAT	1622
Val Thr Thr Ala Val Arg Gly Thr Ile Gly His Ile Ala Pro Glu Tyr	
465 470 475	
CTC TCA ACC GGA AAA TCT TCA GAG AAA ACC GAC GTT TTC GGA TAC GGA	1670
Leu Ser Thr Gly Lys Ser Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly	
480 485 490	
ATC ATG CTT CTA GAA CTA ATC ACA GGA CAA AGA GCT TTC GAT CTC GCT	1718
Ile Met Leu Leu Glu Leu Ile Thr Gly Gln Arg Ala Phe Asp Leu Ala	
495 500 505	
CGG CTA GCT AAC GAC GAC GAC GTC ATG TTA CTT GAC TGG GTG AAA GGA	1766
Arg Leu Ala Asn Asp Asp Asp Val Met Leu Leu Asp Trp Val Lys Gly	
510 515 520	
TTG TTG AAG GAG AAG AAG CTA GAG ATG TTA GTG GAT CCA GAT CTT CAA	1814
Leu Leu Lys Glu Lys Lys Leu Glu Met Leu Val Asp Pro Asp Leu Gln	
525 530 535 540	
ACA AAC TAC GAG GAG AGA GAA CTG GAA CAA GTG ATA CAA GTG GCG TTG	1862
Thr Asn Tyr Glu Glu Arg Glu Leu Glu Gln Val Ile Gln Val Ala Leu	
545 550 555	

00100700 115000

CTA TGC ACG CAA GGA TCA CCA ATG GAA AGA CCA AAG ATG TCT GAA GTT 1910
Leu Cys Thr Gln Gly Ser Pro Met Glu Arg Pro Lys Met Ser Glu Val
560 565 570

GTA AGG ATG CTG GAA GGA GAT GGG CTT GCG GAG AAA TGG GAC GAA TGG 1958
Val Arg Met Leu Glu Gly Asp Gly Leu Ala Glu Lys Trp Asp Glu Trp
575 580 585

CAA AAA GTT GAG ATT TTG AGG GAA GAG ATT GAT TTG AGT CCT AAT CCT 2006
Gln Lys Val Glu Ile Leu Arg Glu Glu Ile Asp Leu Ser Pro Asn Pro
590 595 600

AAC TCT GAT TGG ATT CTT GAT TCT ACT TAC AAT TTG CAC GCC GTT GAG 2054
Asn Ser Asp Trp Ile Leu Asp Ser Thr Tyr Asn Leu His Ala Val Glu
605 610 615 620

TTA TCT GGT CCA AGG TAAAAAAAAA AAAAAAAAAA 2089
Leu Ser Gly Pro Arg
625

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 625 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Met Glu Ser Ser Tyr Val Val Phe Ile Leu Leu Ser Leu Ile Leu Leu
1 5 10 15
Pro Asn His Ser Leu Trp Leu Ala Ser Ala Asn Leu Glu Gly Asp Ala
20 25 30

00100700-11500

Pro Cys Pro Gly Ser Pro Pro Phe Ser Pro Pro Pro Pro Phe Ile Gln
210 215 220

Leu His Asp His Cys Asp Pro Lys Ile Ile His Arg Asp Val Lys Ala

420	425	430
Ala Asn Ile Leu Leu Asp Glu Glu Phe Glu Ala Val Val Gly Asp Phe		
435	440	445
Gly Leu Ala Lys Leu Met Asp Tyr Lys Asp Thr His Val Thr Thr Ala		
450	455	460
Val Arg Gly Thr Ile Gly His Ile Ala Pro Glu Tyr Leu Ser Thr Gly		
465	470	475 480
Lys Ser Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly Ile Met Leu Leu		
485	490	495
Glu Leu Ile Thr Gly Gln Arg Ala Phe Asp Leu Ala Arg Leu Ala Asn		
500	505	510
Asp Asp Asp Val Met Leu Leu Asp Trp Val Lys Gly Leu Leu Lys Glu		
515	520	525
Lys Lys Leu Glu Met Leu Val Asp Pro Asp Leu Gln Thr Asn Tyr Glu		
530	535	540
Glu Arg Glu Leu Glu Gln Val Ile Gln Val Ala Leu Leu Cys Thr Gln		
545	550	555 560
Gly Ser Pro Met Glu Arg Pro Lys Met Ser Glu Val Val Arg Met Leu		
565	570	575
Glu Gly Asp Gly Leu Ala Glu Lys Trp Asp Glu Trp Gln Lys Val Glu		
580	585	590
Ile Leu Arg Glu Glu Ile Asp Leu Ser Pro Asn Pro Asn Ser Asp Trp		
595	600	605
Ile Leu Asp Ser Thr Tyr Asn Leu His Ala Val Glu Leu Ser Gly Pro		
610	615	620

00180700-11500

Arg
625

09180799 111599
86977 86208700